NOT ALL OBSERVATIONS ARE EQUAL: GUIDING DERIVATIVE-FREE ENSEMBLES VIA UNCERTAINTY CONTRACTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Ensemble Kalman Guidance (EnKG) is attractive for high-dimensional inverse problems, especially with black-box physical forward models where repeated gradient computation is difficult or impractical, as it only requires forward evaluations. However, its empirical performance can be hindered by the finite ensemble size when the parameter dimension is large, resulting in unreliable or biased estimates. To address this issue, we propose Uncertainty-Contraction Ensemble Guidance (UCEG), a monitor-guided adaptive sampling framework that improves the quality of the ensemble correction without increasing the ensemble size. UCEG treats guidance construction as an observation-selection problem and uses update activity as a practical proxy for local uncertainty contraction in the finite-ensemble correction, retaining informative observations while suppressing uninformative ones. Experiments on challenging inverse problems including full waveform inversion (FWI) and Navier–Stokes show that UCEG can outperform the corresponding full-observation correction using fewer observations, with faster inference when observation subsampling reduces evaluation cost.

1 INTRODUCTION

Inverse problems arise throughout science and engineering, especially in PDE-governed systems where unknown physical states or parameters are inferred from indirect and noisy observations. Representative examples include full waveform inversion (FWI) (Louboutin et al., 2018; Witte et al., 2018), Navier-Stokes inverse problems (Zheng et al., 2025b), and Darcy flow inverse problems (Yao et al., 2025; Zheng et al., 2025b). In these tasks, data consistency is defined through a physical forward model, whose evaluation can already be computationally demanding. Beyond these benchmarks, many realistic forward systems are available primarily as simulators: they can be queried, but repeated gradient or adjoint computations may be unavailable, unstable, or too expensive for iterative inference. This makes ensemble Kalman methods attractive, since they require only forward evaluations and use empirical covariances to construct derivative-free corrections from an ensemble of candidate states (Evensen, 2003; Iglesias et al., 2013; Zheng et al., 2024). They are therefore well-suited to black-box or expensive simulators where direct differentiation through the forward model is unreliable or unavailable.

Meanwhile, diffusion models offer a complementary advantage for inverse problems. They provide expressive priors for complex high-dimensional fields (Ho et al., 2020; Song et al., 2021; Zheng et al., 2025b), and can be reused across different observation operators without retraining the prior. A central question is then how to inject measurement information during reverse sampling. Gradient-based methods such as DPS use likelihood gradients from the forward model to guide the diffusion trajectory (Chung et al., 2023). This is effective when such gradients are available, but it is less suitable for black-box simulators. To address this setting, Zheng et al. (2024) propose Ensemble Kalman Guidance (EnKG), which constructs the measurement correction from an ensemble of black-box forward evaluations instead of differentiating through the simulator.

However, EnKG still inherits the finite-ensemble limitations of ensemble Kalman methods. In high-dimensional inverse problems, limited ensembles can lead to rank limitations and noisy empirical covariance estimates, motivating dropout and localization strategies (Liu et al., 2025; Tong, 2018;

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Tong & Morzfeld, 2023). We take a complementary observation-side view: for a fixed ensemble size, the quality of the EnKG correction depends on which observation channels enter the empirical update. When the number of channels is large, using all observations is not necessarily beneficial. Redundant or weakly informative channels can amplify finite-ensemble noise and spurious correlations, while a selected subset can produce a more reliable guidance direction.

We therefore cast guidance construction as an observation-selection problem. For a fixed finite ensemble, the main objective is to choose observation units that lead to a more reliable subset-conditioned EnKG correction. This changes the observation side of the empirical update, rather than increasing the ensemble size or modifying the pretrained diffusion prior. In settings where observation units correspond to expensive simulator queries, such as source or receiver selection in FWI, the same selection can also significantly reduce forward-evaluation cost.

To address this observation-selection problem, we propose Uncertainty-Contraction Ensemble Guidance (UCEG), which improves finite-ensemble guidance by adaptively selecting the observations used to form each correction. UCEG is motivated by a local linear-Gaussian identity linking uncertainty contraction to the expected energy of the correction induced by a selected subset. This connection suggests that the realized update activity of previous ensemble corrections can serve as a practical signal for identifying informative observation units.

At each reverse step, UCEG converts the recorded update-activity map into a task-dependent monitor function, and then maps this spatial monitor to sampling scores over observation units. Under a fixed observation budget, most units are sampled according to this monitor, so the next correction focuses on regions where previous updates indicate strong uncertainty contraction. We additionally keep a portion of uniform coverage and uncertainty-based exploration to avoid overly concentrated selection. The resulting subset is used to construct the next subset-conditioned ensemble correction, allowing UCEG to allocate the observation budget to informative units rather than using all observations indiscriminately.

Our contributions are summarized as follows:

- We introduce an observation-side perspective on finite-ensemble guidance: instead of increasing the ensemble size or modifying the pretrained prior, we improve the correction by selecting which observation units enter the empirical update.
- We derive an uncertainty-contraction criterion for subset selection under a local linear-Gaussian conditioning model, and show that the expected energy of the induced mean update matches the corresponding covariance contraction.
- We propose UCEG, which uses realized ensemble update activity to identify informative observations and construct stronger subset-conditioned guidance.
- We validate UCEG on FWI, Navier–Stokes, and Darcy flow inverse problems, showing consistent gains over full-observation EnKG and strong performance against derivative-free baselines under feasible computational budgets.

2 Related works

Ensemble Kalman methods for inverse problems. Ensemble Kalman methods were originally developed for data assimilation (Evensen, 1994; 2003) and later adapted to inverse problems through Ensemble Kalman Inversion (EKI)(Iglesias et al., 2013). These methods are attractive for scientific inverse problems because they construct derivative-free updates using only forward evaluations and empirical ensemble statistics. However, finite ensemble size is a central limitation in high-dimensional settings. The update is restricted by the ensemble subspace, and empirical covariance estimates can be low-rank, noisy, or affected by spurious correlations (Iglesias et al., 2013; Schillings & Stuart, 2017). Existing work addresses these issues through regularization, dropout, and localization strategies (Chada et al., 2020; Liu et al., 2025; Tong, 2018; Tong & Morzfeld, 2023). UCEG is motivated by the same finite-ensemble difficulty, but takes a complementary observation-side perspective: for a fixed ensemble size, we improve the update by selecting which observation units enter the empirical correction.

Diffusion guidance and EnKG. Diffusion models provide expressive priors for complex high-dimensional fields(Ho et al., 2020; Song et al., 2021; Yao et al., 2025) and have been widely used

as plug-and-play priors for inverse problems, where measurement information is injected during sampling without retraining the prior for each forward model (Chung et al., 2023; Song et al., 2023; Zhu et al., 2023; Zheng et al., 2025b). Many guidance methods rely on likelihood gradients, pseudo-inverses, or other structured access to the forward operator. This can be restrictive for black-box scientific simulators. To address this setting, Zheng et al. (2025a) propose Ensemble Kalman Guidance (EnKG), which constructs a derivative-free measurement correction from an ensemble of diffusion particles and their predicted observations. Our work builds on this finite-ensemble correction, but studies a different question: instead of always using the full observation set, we ask which observation units should be selected to form a more reliable subset-conditioned correction.

Observation selection and monitor-based allocation. Observation selection is closely related to optimal experimental design and sensor placement for inverse problems. Bayesian design often values measurements by expected information gain (Huan & Marzouk, 2013), while sensor-placement and Bayesian inverse-problem formulations also use posterior-uncertainty objectives such as A- and D-optimality, based on trace- or determinant-type functions of the posterior covariance (Alexanderian, 2021; Alexanderian et al., 2024). Recent scalable BOED methods further study efficient design criteria for expensive parameter-to-observable maps (Go & Chen, 2025). These objectives provide principled measures of informativeness, but direct optimization typically requires repeatedly evaluating candidate designs through forward solves, adjoint-based sensitivities, posterior sampling, or covariance approximations. This cost makes them difficult to use inside every reverse step of an iterative diffusion or ensemble-guidance sampler.

Monitor-based allocation is a standard idea in adaptive discretization. Classical moving-mesh methods use monitor functions to redistribute mesh nodes toward regions with large solution variation or interpolation error (Huang & Russell, 2010). Recent neural mesh-mover methods extend this idea by learning or constructing mesh movement maps for PDE solvers (Song et al., 2022; Hu et al., 2023). In UCEG, the monitor is not used to move mesh nodes; instead, it converts update activity into observation-unit sampling scores for the next finite-ensemble correction.

3 METHOD

Starting from the premise that finite-ensemble guidance depends on which observation units enter the empirical correction, UCEG formulates EnKG guidance as a budgeted observation-selection problem. Section 3.1 defines the setting and subset-conditioned ensemble corrections; Section 3.2 introduces the uncertainty-contraction target; Section 3.3 connects this target to realized update activity; and Section 3.4 turns this signal into a practical selection rule.

3.1 BLACK-BOX INVERSE PROBLEMS AND LOCAL FINITE-ENSEMBLE CORRECTIONS

Black-box inverse problems. We consider inverse problems whose observations are organized into M observation units,

$$y_q = \mathcal{F}_q(x) + \varepsilon_q, \quad q \in [M], \qquad [M] := \{1, \dots, M\}. \tag{1}$$

Here $x \in \mathbb{R}^d$ is the unknown field, $\mathcal{F}_q$ is the forward operator for the q-th observation unit, and ε_q is the observation noise for unit q. Depending on the application, an observation unit may be a source, receiver, source–receiver block, sensor location, spatial point, or time window. Given an observation budget $K \le M$, we consider the feasible subset space $\mathfrak{I}_K := \{\mathcal{I} \subseteq [M] : |\mathcal{I}| = K\}$. Each $\mathcal{I} \in \mathfrak{I}_K$ denotes one admissible selection of observation units, with restricted observation and forward query

$$y_{\mathcal{I}} := \{y_q\}_{q\in\mathcal{I}}, \qquad \mathcal{F}_{\mathcal{I}}(x) := \{\mathcal{F}_q(x)\}_{q\in\mathcal{I}}.$$

Given the observation $y = \{y_q\}_{q\in[M]}$, we solve the inverse problem using Bayes' rule, $p(x \mid y) \propto p(y \mid x)p(x)$.

Subset-conditioned EnKG correction. Following EnKG (Zheng et al., 2025a), UCEG constructs a derivative-free empirical Kalman correction from particle–observation pairs along the reverse diffusion trajectory. At correction step i, for a selected subset $\mathcal{I} \in \mathfrak{I}_K$, each particle $x_i'^{(j)}$ is paired with the subset-restricted predicted observation $z_{i,\mathcal{I}}^{(j)} := \mathcal{F}_{\mathcal{I}}(\phi_i(x_i'^{(j)}))$, where ϕ_i maps the current

reverse-time particle to its clean endpoint. Let

$$\bar{x}'_i := \frac{1}{J}\sum_{j=1}^{J} x_i'^{(j)}, \qquad \bar{z}_{i,\mathcal{I}} := \frac{1}{J}\sum_{j=1}^{J} z_{i,\mathcal{I}}^{(j)},$$

where J denotes the ensemble size. The subset-conditioned EnKG direction and particle update are

$$g_{i,\mathcal{I}}^{(j)} := \frac{1}{J}\sum_{k=1}^{J} \left\langle z_{i,\mathcal{I}}^{(k)} - \bar{z}_{i,\mathcal{I}}, y_{\mathcal{I}} - z_{i,\mathcal{I}}^{(j)} \right\rangle_{R_{\mathcal{I}}} \left(x_i'^{(k)} - \bar{x}'_i \right), \qquad x_{i+1}^{(j)} = x_i'^{(j)} + w_i g_{i,\mathcal{I}}^{(j)}, \tag{2}$$

where $\langle a, b\rangle_{R_{\mathcal{I}}} := a^\top R_{\mathcal{I}}^{-1} b$, $R_{\mathcal{I}}$ is the observation-noise covariance restricted to $\mathcal{I}$, and w_i is the guidance weight. Eq. equation 2 shows that the observation subset directly determines the EnKG correction. It selects the residual $y_{\mathcal{I}} - z_{i,\mathcal{I}}^{(j)}$ and also determines which predicted observation features are used to form the particle–observation correlations. When M is large and the ensemble size J is limited, using all observation units can make these empirical correlations less reliable, due to redundant measurements and finite-ensemble noise. Therefore, instead of always using the full observation set, UCEG selects a budgeted subset before forming the correction. At reverse step i, given budget K, we seek

$$\mathcal{I}_i^\star = \arg\max_{\mathcal{I}\in\mathfrak{I}_K} U_i(\mathcal{I}), \tag{3}$$

where $U_i(\mathcal{I})$ measures the usefulness of subset $\mathcal{I}$ for the current correction. The next subsections define this utility via uncertainty contraction and connect it to a practical update-activity monitor.

3.2 Uncertainty-contraction utility

We next define the subset utility $U_i(\mathcal{I})$ in Eq. equation 3. EnKG derives its derivative-free correction by leveraging statistical linearization from ensemble Kalman methods. Following this perspective, we analyze an idealized local linear-Gaussian conditioning model in which the uncertainty reduction induced by a candidate subset can be characterized analytically:

$$x'_i \sim \mathcal{N}(\bar{x}'_i, C_{xx}^{(i)}), \qquad y_{\mathcal{I}} = A_{i,\mathcal{I}} x'_i + b_{i,\mathcal{I}} + \eta_{\mathcal{I}}, \qquad \eta_{\mathcal{I}} \sim \mathcal{N}(0, R_{\mathcal{I}}). \tag{4}$$

Here x'_i denotes a generic predicted reverse-time particle at step i, and $\bar{x}'_i$, $C_{xx}^{(i)}$ are the empirical mean and covariance of the predicted ensemble. The pair $(A_{i,\mathcal{I}}, b_{i,\mathcal{I}})$ specifies the linearized observation relation for the selected subset $\mathcal{I}$; only $A_{i,\mathcal{I}}$ enters the covariance update.

With the innovation covariance $S_{i,\mathcal{I}} := A_{i,\mathcal{I}} C_{xx}^{(i)} A_{i,\mathcal{I}}^\top + R_{\mathcal{I}}$, Gaussian conditioning gives the posterior covariance and covariance contraction as:

$$C_{xx}^{(i),+}(\mathcal{I}) = C_{xx}^{(i)} - C_{xx}^{(i)} A_{i,\mathcal{I}}^\top S_{i,\mathcal{I}}^{-1} A_{i,\mathcal{I}} C_{xx}^{(i)}, \qquad \Delta C_{xx}^{(i)}(\mathcal{I}) := C_{xx}^{(i)} - C_{xx}^{(i),+}(\mathcal{I}). \tag{5}$$

We define the ideal one-step utility as the total covariance contraction,

$$U_i(\mathcal{I}) := \operatorname{tr}\left(\Delta C_{xx}^{(i)}(\mathcal{I}) \right). \tag{6}$$

This utility ranks subsets by the total reduction in the current state covariance after conditioning. Under the local linear-Gaussian model, a larger value therefore indicates stronger expected one-step uncertainty contraction. We use this utility as an idealized analytical reference for subset selection, rather than assuming that the practical nonlinear finite-ensemble correction itself follows the Gaussian model.

For FWI, $\mathcal{I}$ has a concrete acquisition meaning. Let m denote the unknown medium model. For source s, the forward wavefield is

$$u_s(m) = \mathcal{A}(m)^{-1} P_s^\top q_s, \qquad d_{s,r}(m) = P_r u_s(m),$$

where $\mathcal{A}(m)$ is the discretized wave-equation operator, $P_s^\top$ injects the source, and P_r samples the wavefield at receiver r (Louboutin et al., 2018; Witte et al., 2018). An observation unit may be a source, a receiver, or a source–receiver block. Given a selected set of units $\mathcal{I}$, let $\Pi(\mathcal{I})$ denote the active source–receiver pairs included by this selection.The restricted FWI forward map is therefore

$$\mathcal{F}_{\mathcal{I}}(m) = \{d_{s,r}(m)\}_{(s,r)\in\Pi(\mathcal{I})} = \left\{ P_r \mathcal{A}(m)^{-1} P_s^\top q_s \right\}_{(s,r)\in\Pi(\mathcal{I})}.$$

Thus, in FWI, $U_i(\mathcal{I})$ ranks acquisition subsets by the one-step uncertainty reduction they induce under the local conditioning model.

Eq. equation 6 gives a principled one-step target for subset selection: an informative subset should maximize contraction of the current state uncertainty. The next subsection shows that this target can be monitored in practice through realized EnKG update activity.

3.3 Update activity as a proxy for uncertainty contraction

The previous subsection establishes an exact uncertainty-contraction identity under an idealized local linear-Gaussian conditioning model. This identity does not generally hold exactly for the practical nonlinear finite-ensemble EnKG correction. We therefore interpret realized update activity as a ranking proxy rather than a numerical estimator of the ideal utility. The theoretical result links update energy to uncertainty contraction, while Appendix D empirically validates the proxy in the practical nonlinear setting.

Under the same model, define

$$S_{i,\mathcal{I}} := A_{i,\mathcal{I}} C_{xx}^{(i)} A_{i,\mathcal{I}}^\top + R_{\mathcal{I}}, \qquad K_{i,\mathcal{I}} := C_{xx}^{(i)} A_{i,\mathcal{I}}^\top S_{i,\mathcal{I}}^{-1}.$$

Gaussian conditioning gives

$$\bar{x}_i^+(\mathcal{I}) = \bar{x}_i' + K_{i,\mathcal{I}} \left(y_{\mathcal{I}} - A_{i,\mathcal{I}} \bar{x}_i' - b_{i,\mathcal{I}} \right),$$

together with the covariance update in Eq. equation 5.

Proposition 1 (Mean-update energy equals covariance contraction)**.** *Under the linear-Gaussian model in Eq. equation 4, define the subset-induced mean update*

$$\Delta \bar{x}_i(\mathcal{I}) := \bar{x}_i^+(\mathcal{I}) - \bar{x}_i'. \tag{7}$$

Then

$$\mathbb{E}\left[\Delta \bar{x}_i(\mathcal{I}) \Delta \bar{x}_i(\mathcal{I})^\top \right] = \Delta C_{xx}^{(i)}(\mathcal{I}), \tag{8}$$

where the expectation is taken over the predictive distribution of $y_{\mathcal{I}}$ induced by Eq. equation 4.

The derivation is provided in Appendix A. Proposition 1 shows that, in the linear-Gaussian model of Eq. equation 4, the second moment of the subset-induced mean update is exactly the covariance contraction caused by conditioning on $y_{\mathcal{I}}$. Equivalently, for any linear projection L_r onto a spatial location, block, or component,

$$\mathbb{E}\left[\| L_r \Delta \bar{x}_i(\mathcal{I}) \|_2^2 \right] = \operatorname{tr}\left[L_r \Delta C_{xx}^{(i)}(\mathcal{I}) L_r^\top \right]. \tag{9}$$

Thus, in this model, larger expected projected update energy corresponds to larger projected covariance contraction. This identity motivates the monitor used in the EnKG sampler. After the EnKG correction at reverse step i, we record the realized particle displacement

$$\delta_i^{(j)} := x_{i+1}^{(j)} - x_i'^{(j)}, \tag{10}$$

where $x_i'^{(j)}$ is the predicted particle before correction and $x_{i+1}^{(j)}$ is the corrected particle. For spatially structured states, we define the update-activity map

$$H_i(r) := \frac{1}{J} \sum_{j=1}^{J} \left\| \delta_i^{(j)}(r) \right\|_2. \tag{11}$$

Large values of $H_i(r)$ indicate locations where the previous EnKG correction acted strongly. Guided by Proposition 1, we use this realized update activity as a computationally cheap signal for selecting observations in the next correction step.

3.4 Monitor-guided observation selection

We use the update-activity signal from Section 3.3 to construct a spatial monitor for observation selection. This follows the monitor-function principle used in adaptive discretization, where a monitor specifies where sampling density should be increased (Hu et al., 2023). Here, the monitor is used

to select observation units for the subset-restricted EnKG correction. If previous EnKG corrections repeatedly move the state in a region, then observations associated with that region are expected to produce larger uncertainty contraction in subsequent corrections.

At reverse step i, we define the monitor on the target-state grid Ω_x, where the update-activity map H_{i-1} is recorded. This grid corresponds to the unknown field domain, such as the velocity-model grid in FWI. We construct the monitor map by normalizing the cached update activity on Ω_x and applying a task-dependent monitor function:

$$\bar{H}_{i-1}(r) := \frac{H_{i-1}(r)}{|\Omega_x|^{-1}\sum_{r'\in\Omega_x} H_{i-1}(r') + \epsilon_0}, \qquad m_i(r) := \mathcal{M}(\bar{H}_{i-1})(r), \qquad r \in \Omega_x. \tag{12}$$

The choice of $\mathcal{M}$ is flexible and can be adapted to the task. Following the link between update activity and uncertainty contraction in Section 3.3, we use the identity map in the main experiments, so $m_i = \bar{H}_{i-1}$. Next, we map the target-state monitor m_i to the observation-unit space. Let $\mathcal{Q} = [M]$ be the set of candidate observation units, and let $\omega_q(r) \geq 0$ denote the projection kernel from target-state location $r \in \Omega_x$ to unit q. We define

$$a_i(q) := \mathcal{P}_{\mathcal{Q}}[m_i](q) = \frac{\sum_{r\in\Omega_x} \omega_q(r) m_i(r)}{\sum_{r\in\Omega_x} \omega_q(r) + \epsilon_0}, \qquad q \in \mathcal{Q}. \tag{13}$$

For spatial observations, ω_q reduces to a point-selection or local block-selection kernel. For FWI, ω_q projects the target-state monitor to source or receiver scores through the acquisition geometry. The projection produces a unit-score vector $a_i \in \mathbb{R}^M$. We normalize it as$\pi_i(q) = a_i(q)/(\sum_{\tilde{q}\in\mathcal{Q}} a_i(\tilde{q}) + \varepsilon)$, and sample distinct observation units from $\mathcal{Q}$ without replacement according to π_i. In this case, units with larger projected monitor values are more likely to be included in the next subset-restricted EnKG correction.

In practice, UCEG is controlled by three quantities: the observation budget, effective uniform ratio, and uncertainty extra budget. The observation budget sets the number of main selected units; within it, the effective uniform ratio controls uniform coverage, and the remaining units are sampled from the monitor distribution π_i. This prevents the subset from concentrating only on high-update-activity regions. The uncertainty extra budget adds a small number of exploratory units outside the main budget. We compute $V_i(r) := \mathrm{std}_j(\hat{x}^{(j)}_{N,i}(r))$, with $\hat{x}^{(j)}_{N,i} = \phi_i(x'^{(j)}_i)$, project V_i using Eq. equation 13, and sample additional units from the remaining candidates. The final subset combines uniform-coverage, monitor-sampled, and uncertainty-extra units to form $z^{(j)}_{i,\mathcal{I}_i} := \mathcal{F}_{\mathcal{I}_i}(\phi_i(x'^{(j)}_i))$ for the EnKG correction.

4 Experiments

To evaluate whether observation selection improves finite-ensemble guidance, we test UCEG on FWI, Navier-Stokes, and Darcy flow inverse problems. The experimental setting is described in Section 4.1. Section 4.2 compares UCEG with EnKG and other baselines to evaluate reconstruction quality and computational efficiency. Finally, Section 4.3 studies the main selection components, including receiver budget, uncertainty extra budget, and effective uniform ratio.

4.1 Experimental Setting

Tasks and data. We evaluate UCEG on three PDE inverse problems: full waveform inversion (FWI), Navier-Stokes, and Darcy flow. For FWI and Navier-Stokes, we follow the InverseBench evaluation protocols (Zheng et al., 2025b) where each task has 10 test cases. For Darcy flow, we use the GRF/lognormal-coefficient model and FEM discretization described in Appendix B.3. We train a FunDPS prior (Yao et al., 2025) on 100,000 Darcy samples and evaluate inference on 20 held-out test cases.

These three tasks cover complementary PDE inverse-problem settings: wave-based imaging, nonlinear fluid dynamics, and elliptic PDE parameter estimation. They also differ in their observation units: FWI uses acquisition units such as sources and receivers, while Navier-Stokes and Darcy flow use spatial measurement locations. Detailed task definitions, observation models, and evaluation metrics are provided in Appendix B.

Table 1: FWI main results. Mean and standard deviation are reported over completed cases. Runtime is average observation plus inference wall time per completed case.

Method	Rel. L2↓	PSNR↑	SSIM↑	Time / case
DPG	0.400 (0.055)	8.204 (0.702)	0.100 (0.020)	3.4 ± 0.1 h
DPS	0.220 (0.034)	13.193 (1.362)	0.122 (0.022)	3.5 ± 0.7 min
SCG	0.333 (0.076)	10.290 (1.705)	0.242 (0.053)	2.1 ± 0.3 h
DPS-cGSG	0.245 (0.027)	12.240 (1.162)	0.134 (0.019)	4.7 ± 0.1 h
DPS-fGSG	0.241 (0.036)	12.388 (1.471)	0.212 (0.039)	5.0 ± 0.3 h
EKI	0.053 (0.007)	10.437 (0.965)	0.386 (0.059)	46.7 ± 1.3 min
EnKG	0.026 (0.010)	14.301 (2.974)	0.398 (0.117)	7.7 ± 0.3 h
UCEG(random)	0.024 (0.007)	15.063 (2.431)	0.472 (0.075)	32.3 ± 3.5 min
UCEG(uniform)	0.023 (0.007)	15.062 (2.087)	0.479 (0.068)	33.4 ± 1.6 min
UCEG(random monitor)	0.022 (0.005)	15.205 (1.441)	0.476 (0.061)	31.4 ± 0.5 min
UCEG	**0.021** (0.004)	**15.432** (1.412)	**0.481** (0.061)	39.2 ± 1.0 min

Baselines and implementation setup. We compare UCEG with EnKG, EKI, and several diffusion-guidance baselines, including DPG, SCG, DPS, DPS-cGSG, and DPS-fGSG. EnKG denotes the standard full-observation EnKG correction, while EKI is included as a classical derivative-free ensemble baseline. We additionally consider three observation-budget-matched UCEG variants to isolate the effect of adaptive selection. UCEG(random) selects the entire observation subset randomly. UCEG(uniform) uses uniformly distributed observations with the same observation budget. UCEG(random monitor) keeps the complete UCEG pipeline unchanged but replaces only the observations selected by the update-history monitor with random observations. For FWI, several baselines require multiple days for a single test case, so we report computationally feasible stable configurations under our available budget. For Navier–Stokes, we follow the InverseBench protocol (Zheng et al., 2025b). For Darcy flow, we use matched solver, prior, and observation settings across the compared selection strategies. Detailed baseline settings are provided in Appendix C.

UCEG implementation. UCEG uses the subset-conditioned EnKG correction in Section 3.1. At each correction step, UCEG selects an observation subset $\mathcal{I}_i$, forms the subset-restricted predictions $z_{i,\mathcal{I}_i}^{(j)} = \mathcal{F}_{\mathcal{I}_i}(\phi_i(x_i'^{(j)}))$, and constructs the EnKG correction using this selected subset. The selection module uses the update-activity monitor as the main signal, together with uniform coverage and uncertainty-based exploration as described in Section 3.4.

For FWI, unless otherwise specified, UCEG uses 8 selected sources out of 16, time stride 4, receiver budget 20 out of 129, uncertainty extra budget 5%, and effective uniform ratio 25%; The main solver settings are 64 particles, 40 reverse steps, and guidance scale 2.0 for FWI; 2048 particles, 80 reverse steps, and guidance scale 2.0 for Navier–Stokes; and 256 particles, 80 reverse steps, and guidance scale 0.20 for Darcy flow. Wall-clock timings are measured on one NVIDIA A100 GPU.

4.2 Main Results

We first compare UCEG with full-observation EnKG and the broader inverse-problem baselines to evaluate reconstruction quality and computational cost. We then use observation-budget-matched UCEG variants to distinguish the benefit of reducing the observation set from the benefit of how the reduced budget is allocated. UCEG(random) tests whether arbitrary observation reduction is sufficient, while UCEG(uniform) tests whether simple uniform coverage explains the improvement. Most importantly, UCEG(random monitor) differs from UCEG only in the observations assigned by the update-history monitor, and therefore directly isolates the additional contribution of monitor-guided selection.

Full waveform inversion. Table 1 reports the main FWI results. Compared with EnKG, UCEG improves all three reconstruction metrics while substantially reducing runtime: EnKG takes 7.7 hours per case, whereas UCEG takes 39.2 minutes per case. This comparison shows that using a selected acquisition subset can improve the finite-ensemble correction while reducing the acoustic forward evaluations required in FWI. Under the same observation budget, UCEG also outperforms UCEG(random), UCEG(uniform), and UCEG(random monitor), reducing the relative ℓ_2 error by 11.3%, 6.3%, and 3%, respectively. In particular, UCEG(random monitor) differs only in the monitor-selected receivers, directly isolating the additional benefit of the update-history monitor.

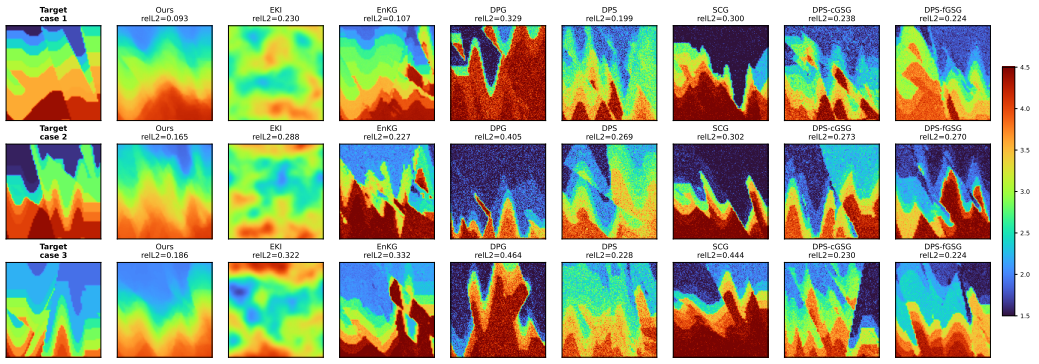

Figure 1: Representative FWI reconstruction results.

Compared with the broader FWI baselines, UCEG achieves the best reconstruction quality among the evaluated methods, with the lowest relative ℓ_2 error and the highest PSNR and SSIM. Gradient-based DPS is fast in the reduced FWI setting but gives substantially weaker reconstructions. Search- or query-based guidance baselines such as DPG, SCG, DPS-cGSG, and DPS-fGSG require substantially larger inference time in this task and still underperform UCEG. These results suggest that, under the feasible computational settings we tested, ensemble-based guidance with adaptive observation selection offers a favorable quality–cost trade-off for FWI. A paired trajectory analysis further shows that the update-history monitor leads to stronger empirical uncertainty contraction than UCEG(random monitor); detailed definitions, statistics, and results are provided in Appendix D.

Navier–Stokes and Darcy flow. Tables 2 and 3 report results on the two additional PDE inverse problems. On Navier–Stokes, UCEG achieves the best relative ℓ_2 error across all observation settings. The improvement over EnKG is consistent under different downsampling factors and noise levels, and the gap becomes more pronounced in the more challenging sparse-observation settings. This indicates that update-activity-based observation selection is not specific to FWI acquisition geometry; it also improves finite-ensemble guidance when the observation units are spatial measurement locations.

On Darcy flow, UCEG consistently improves over EnKG across ensemble sizes under matched solver, prior, measurement model, and preprocessing settings. Together, the Navier–Stokes and Darcy results support UCEG as a general observation-selection mechanism for high-dimensional PDE inverse problems. Additional observation-budget-matched comparisons with UCEG(random) and UCEG(random monitor) are provided in Appendix E.

Qualitative results. Figure 1 shows that UCEG yields velocity fields with clearer spatial continuity and fewer large noisy artifacts than the compared baselines, consistent with the quantitative gains in Table 1. Additional qualitative visualizations are provided in Appendix G, including the corresponding FWI error maps, Navier-Stokes reconstruction and error maps, and a Darcy monitor-selection visualization. The Darcy visualization shows that the monitor progressively concentrates selections on regions where the posterior-mean reconstruction remains inconsistent with the target, illustrating how update activity guides observation selection during inference. To further verify the quality of the reconstructed FWI velocity fields beyond standard pixel-wise metrics, we evaluate both gradient-domain reconstruction accuracy and sensitivity to the number of reduced-observation corrections in Appendix F.

4.3 Ablation Study

We conduct three FWI ablations to study the receiver-side choices in UCEG: receiver budget, uncertainty extra budget, and effective uniform ratio. Unless otherwise specified, we use the default FWI setting with 8 selected sources out of 16, time stride 4, receiver budget 20 out of 129, uncertainty extra budget 5%, and effective uniform ratio 25%. Results are averaged over 10 cases and reported in Table 4.

Receiver budget. We vary the receiver budget from 20 to 80. UCEG remains stable across all budgets: relative ℓ_2 stays around 0.022, and PSNR/SSIM change only slightly. This suggests that, once UCEG preserves its selection rule, increasing the receiver budget is not the dominant factor; a compact selected receiver subset is already sufficient for effective correction.

Table 2: Navier–Stokes inverse-problem results under multiple observation settings.

Method	$\times 2,\ \sigma = 0$	$\times 2,\ \sigma = 2$	$\times 4,\ \sigma = 2$	$\times 8,\ \sigma = 2$
EKI	0.577 (0.138)	0.673 (0.107)	0.805 (0.112)	1.116 (0.090)
DPS-fGSG	1.687 (0.156)	1.454 (0.154)	1.200 (0.100)	1.260 (0.117)
DPS-cGSG	2.03 (0.314)	1.746 (0.191)	1.114 (0.144)	1.218 (0.113)
DPG	0.325 (0.188)	0.466 (0.171)	0.454 (0.207)	0.846 (0.251)
SCG	0.908 (0.600)	0.966 (0.546)	0.929 (0.505)	1.347 (0.141)
EnKG	0.120 (0.085)	0.339 (0.049)	0.619 (0.100)	0.801 (0.201)
UCEG	**0.104 (0.044)**	**0.250 (0.047)**	**0.373 (0.077)**	**0.541 (0.207)**

Table 3: Darcy reconstruction on the target g channel under matched EnKG settings.

N	EnKG $\ell_2 \downarrow$	UCEG $\ell_2 \downarrow$	Gain	EnKG min/sample	UCEG min/sample
64	0.1600	**0.1595**	0.0005	4.68	**4.67**
128	0.1451	**0.1443**	0.0008	9.24	**9.18**
256	0.1230	**0.1206**	0.0024	19.07	**18.40**

Table 4: FWI ablations for our method. Source setting is fixed to the current best setup.

Study	Metric	Setting				
Receiver budget		20	40	50	60	80
	Rel. L2↓	**0.0217**±0.0042	0.0220±0.0042	0.0218±0.0041	0.0221±0.0042	0.0221±0.0043
	PSNR↑	**15.294**±1.491	15.185±1.429	15.268±1.495	15.115±1.408	15.151±1.439
	SSIM↑	0.479±0.060	0.478±0.060	**0.480**±0.061	0.478±0.062	0.476±0.059
Uncertainty budget		0%	2%	5%	10%	20%
	Rel. L2↓	0.0231±0.0086	0.0217±0.0042	**0.0217**±0.0042	0.0217±0.0042	0.0217±0.0042
	PSNR↑	15.050±2.383	15.288±1.490	**15.294**±1.491	15.283±1.479	15.280±1.497
	SSIM↑	**0.484**±0.078	0.478±0.060	0.479±0.060	0.479±0.059	0.478±0.059
Uniform / exploration		25%	50%	90%		
	Rel. L2↓	**0.0213**±0.0040	0.0218±0.0040	0.0220±0.0041		
	PSNR↑	**15.432**±1.412	15.259±1.462	15.151±1.435		
	SSIM↑	**0.481**±0.061	0.479±0.060	0.477±0.062		

Uncertainty extra budget. We then fix the receiver budget to 20 and vary the uncertainty extra budget from 0% to 20%. Compared with using no uncertainty extra budget, adding a small uncertainty budget improves relative ℓ_2 and PSNR compared with using no uncertainty extra budget. The gains exhibit diminishing returns: 5% gives the best PSNR and matches the best relative ℓ_2, while larger values provide no clear improvement and may introduce less useful observations.

Effective uniform ratio. We fix the receiver budget to 20 and the uncertainty extra budget to 5%, and vary the effective uniform ratio from 25% to 90%. Larger uniform ratios slightly degrade performance, indicating that moderate uniform coverage helps avoid over-concentration in high-update-activity regions, whereas excessive uniform sampling weakens the UCEG selection effect. These ablations show that preserving the UCEG selection rule is more important than simply increasing the receiver budget. Under a fixed budget, excessive uncertainty-driven or uniform exploration can dilute update-activity-based selection and slightly reduce performance.

5 Conclusion

We proposed UCEG, an adaptive observation-selection method for finite-ensemble diffusion guidance in black-box inverse problems. UCEG formalizes subset selection through a local uncertainty-contraction utility and uses realized ensemble update activity as a monitor for selecting informative observations. By constructing each correction from a selected subset rather than the full observation set, UCEG improves the reliability of finite-ensemble guidance. Across FWI, Navier–Stokes, and Darcy flow, UCEG shows consistent improvements over EnKG. The gain is especially clear in FWI, where source and receiver selection improves reconstruction quality and substantially reduces inference time.

AI USE STATEMENT

In this work, we used OpenAI GPT-6 for language refinement, including correcting grammatical errors and improving the overall readability of the manuscript, as well as for software-development assistance, including debugging and optimizing the experimental code implementation. All AI-assisted text and code suggestions were reviewed and verified by the authors before inclusion. The authors take full responsibility for the final content of this work, including the manuscript, code, and reported experimental results.

ETHICS STATEMENT

This work does not involve human subjects, personally identifiable information, or sensitive user data. The experiments are conducted using publicly available benchmark resources and synthetically generated PDE data, and we follow the corresponding licenses and terms of use. UCEG is intended as a methodological contribution for scientific inverse problems and does not directly make decisions about individuals. We are not aware of additional ethical concerns specific to this work.

REPRODUCIBILITY STATEMENT

Detailed instructions for dataset access, data preprocessing, experimental settings, and method implementation are provided in Section 4 and the Appendix. The Appendix further specifies the task configurations, baseline and UCEG hyperparameters, observation-selection settings, and complete derivation of the theoretical result. We use publicly available benchmark protocols and pretrained priors from InverseBench (Zheng et al., 2025b) and FunDPS (Yao et al., 2025), following the corresponding licenses and terms of use. Our code and configuration files will be released publicly upon acceptance.

REFERENCES

Alen Alexanderian. Optimal experimental design for infinite-dimensional bayesian inverse problems governed by pdes: A review. *Inverse Problems*, 37(4):043001, 2021. doi: 10.1088/1361-6420/abe10c.

Alen Alexanderian, Ruanui Nicholson, and Noemi Petra. Optimal design of large-scale nonlinear bayesian inverse problems under model uncertainty. *Inverse Problems*, 40(9):095001, 2024.

Neil K Chada, Andrew M Stuart, and Xin T Tong. Tikhonov regularization within ensemble kalman inversion. *SIAM Journal on Numerical Analysis*, 58(2):1263–1294, 2020.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5): 10143–10162, 1994.

Geir Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003. doi: 10.1007/s10236-003-0036-9.

Jinwoo Go and Peng Chen. Accurate, scalable, and efficient bayesian optimal experimental design with derivative-informed neural operators. *Computer Methods in Applied Mechanics and Engineering*, 438:117845, 2025.

Jonathan Ho, Ajay N. Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Peiyan Hu, Yue Wang, and Zhi-Ming Ma. Better neural pde solvers through data-free mesh movers. *arXiv preprint arXiv:2312.05583*, 2023.

Xun Huan and Youssef M Marzouk. Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317, 2013.

Weizhang Huang and Robert D Russell. *Adaptive moving mesh methods*, volume 174. Springer Science & Business Media, 2010.

Marco A. Iglesias, Kody J. H. Law, and Andrew M. Stuart. Ensemble kalman methods for inverse problems. *Inverse Problems*, 29(4):045001, 2013. doi: 10.1088/0266-5611/29/4/045001.

Shuigen Liu, Sebastian Reich, and Xin T Tong. Dropout ensemble kalman inversion for high dimensional inverse problems. *SIAM Journal on Numerical Analysis*, 63(2):685–715, 2025.

Mathias Louboutin, Philipp A. Witte, Michael Lange, Navjot Kukreja, Fabio Luporini, Gerard Gorman, and Felix J. Herrmann. Full-waveform inversion, part 2: Adjoint modeling. *The Leading Edge*, 37(1):69–72, 2018. doi: 10.1190/tle37010069.1.

Claudia Schillings and Andrew M Stuart. Analysis of the ensemble kalman filter for inverse problems. *SIAM Journal on Numerical Analysis*, 55(3):1264–1290, 2017.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International conference on learning representations*, 2023.

Wenbin Song, Mingrui Zhang, Joseph G Wallwork, Junpeng Gao, Zheng Tian, Fanglei Sun, Matthew Piggott, Junqing Chen, Zuoqiang Shi, Xiang Chen, et al. M2n: Mesh movement networks for pde solvers. *Advances in Neural Information Processing Systems*, 35:7199–7210, 2022.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Xin T. Tong. Performance analysis of local ensemble kalman filter. *Journal of Nonlinear Science*, 28 (4):1397–1442, 2018. doi: 10.1007/s00332-018-9453-2.

Xin T. Tong and Matthias Morzfeld. Localized ensemble kalman inversion. *Inverse Problems*, 39(6): 064002, 2023. doi: 10.1088/1361-6420/accb08.

Philipp A. Witte, Mathias Louboutin, Keegan Lensink, Michael Lange, Navjot Kukreja, Fabio Luporini, Gerard Gorman, and Felix J. Herrmann. Full-waveform inversion, part 3: Optimization. *The Leading Edge*, 37(2):142–145, 2018. doi: 10.1190/tle37020142.1.

Jiachen Yao, Abbas Mammadov, Julius Berner, Gavin Kerrigan, Jong Chul Ye, Kamyar Azizzadenesheli, and Anima Anandkumar. Guided diffusion sampling on function spaces with applications to PDEs. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=oAgwvZay2U.

Hongkai Zheng, Wenda Chu, Austin Wang, Nikola Kovachki, Ricardo Baptista, and Yisong Yue. Ensemble kalman diffusion guidance: A derivative-free method for inverse problems. *arXiv preprint arXiv:2409.20175*, 2024.

Hongkai Zheng, Wenda Chu, Austin Wang, Nikola B. Kovachki, Ricardo Baptista, and Yisong Yue. Ensemble kalman diffusion guidance: A derivative-free method for inverse problems. *Transactions on Machine Learning Research*, 2025a.

Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy T. Feng, Caifeng Zou, Yu Sun, Nikola Kovachki, Zachary E. Ross, Katherine L. Bouman, and Yisong Yue. Inversebench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences. In *International Conference on Learning Representations*, 2025b. Spotlight.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1219–1229, 2023.

A Derivation of the Update-Activity Identity

We use the notation from Section 3.3. Throughout this derivation, the correction step i and the selected observation subset $\mathcal{I}$ are fixed. Under the linear-Gaussian model,

$$x'_i \sim \mathcal{N}(\bar{x}'_i, C^{(i)}_{xx}), \qquad y_{\mathcal{I}} = A_{i,\mathcal{I}} x'_i + \eta_{\mathcal{I}}, \qquad \eta_{\mathcal{I}} \sim \mathcal{N}(0, R_{\mathcal{I}}).$$

We write the innovation covariance and linear conditioning gain as

$$S_{i,\mathcal{I}} = A_{i,\mathcal{I}} C^{(i)}_{xx} A^{\top}_{i,\mathcal{I}} + R_{\mathcal{I}}, \qquad K_{i,\mathcal{I}} = C^{(i)}_{xx} A^{\top}_{i,\mathcal{I}} S^{-1}_{i,\mathcal{I}}.$$

Gaussian conditioning gives

$$\bar{x}^{+}_i(\mathcal{I}) = \bar{x}'_i + K_{i,\mathcal{I}} \left(y_{\mathcal{I}} - A_{i,\mathcal{I}} \bar{x}'_i \right), \qquad C^{(i),+}_{xx}(\mathcal{I}) = C^{(i)}_{xx} - C^{(i)}_{xx} A^{\top}_{i,\mathcal{I}} S^{-1}_{i,\mathcal{I}} A_{i,\mathcal{I}} C^{(i)}_{xx}.$$

Thus

$$\Delta C^{(i)}_{xx}(\mathcal{I}) = C^{(i)}_{xx} - C^{(i),+}_{xx}(\mathcal{I}) = C^{(i)}_{xx} A^{\top}_{i,\mathcal{I}} S^{-1}_{i,\mathcal{I}} A_{i,\mathcal{I}} C^{(i)}_{xx}.$$

Proof. Let

$$\nu_{i,\mathcal{I}} = y_{\mathcal{I}} - A_{i,\mathcal{I}} \bar{x}'_i$$

be the innovation. From the linear-Gaussian model,

$$\nu_{i,\mathcal{I}} = A_{i,\mathcal{I}}(x'_i - \bar{x}'_i) + \eta_{\mathcal{I}}, \qquad \mathbb{E}[\nu_{i,\mathcal{I}} \nu^{\top}_{i,\mathcal{I}}] = S_{i,\mathcal{I}}.$$

The subset-induced mean update is

$$\Delta \bar{x}_i(\mathcal{I}) = \bar{x}^{+}_i(\mathcal{I}) - \bar{x}'_i = K_{i,\mathcal{I}} \nu_{i,\mathcal{I}}.$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\Delta \bar{x}_i(\mathcal{I}) \Delta \bar{x}_i(\mathcal{I})^{\top} \right] &= \mathbb{E}\left[K_{i,\mathcal{I}} \nu_{i,\mathcal{I}} \nu^{\top}_{i,\mathcal{I}} K^{\top}_{i,\mathcal{I}} \right] \\
&= K_{i,\mathcal{I}} S_{i,\mathcal{I}} K^{\top}_{i,\mathcal{I}} \\
&= \left(C^{(i)}_{xx} A^{\top}_{i,\mathcal{I}} S^{-1}_{i,\mathcal{I}} \right) S_{i,\mathcal{I}} \left(S^{-1}_{i,\mathcal{I}} A_{i,\mathcal{I}} C^{(i)}_{xx} \right) \\
&= C^{(i)}_{xx} A^{\top}_{i,\mathcal{I}} S^{-1}_{i,\mathcal{I}} A_{i,\mathcal{I}} C^{(i)}_{xx} \\
&= \Delta C^{(i)}_{xx}(\mathcal{I}). \qquad (14)
\end{aligned}$$

This proves the identity. □

B Dataset Details

B.1 Full Waveform Inversion

We use the FWI evaluation protocol of InverseBench Zheng et al. (2025b). The inverse task is to recover the subsurface compressional velocity field $v(x)$ from seismic waveform measurements recorded at receivers. For a source $q(x,t)$, the pressure wavefield $u(x,t)$ is modeled by the acoustic scalar wave equation

$$\frac{1}{v(x)^2} \frac{\partial^2 u(x,t)}{\partial t^2} - \nabla^2 u(x,t) = q(x,t). \qquad (15)$$

After discretization, we write the wave-equation solve abstractly as

$$A(v)u = q, \qquad (16)$$

where $A(v)$ is the discretized acoustic wave operator. The observed waveform data are obtained by sampling the pressure wavefield at receiver locations,

$$y = Pu + n, \qquad (17)$$

where P denotes receiver sampling and n is measurement noise.

FWI is high-dimensional and computationally expensive because evaluating the forward operator requires wave-equation solves over the seismic acquisition geometry. It is also strongly nonlinear, and conventional optimization-based FWI typically depends on repeated gradient computations and a sufficiently good initial model. In UCEG, observation units are defined by the acquisition geometry: depending on the budget setting, we select sources, receiver groups, or source–receiver blocks. We evaluate on 10 test cases.

B.2 NAVIER–STOKES

We use the Navier–Stokes inverse problem from InverseBench Zheng et al. (2025b). The task is to infer the initial vorticity field from noisy partial observations of the vorticity at a later time. The forward model is the two-dimensional incompressible Navier–Stokes equation in vorticity form on a periodic domain:

$$\partial_t w(x,t) + u(x,t)\cdot\nabla w(x,t) = \nu\Delta w(x,t) + f(x), \qquad x\in(0,2\pi)^2,\ t\in[0,T], \tag{18}$$

$$\nabla\cdot u(x,t) = 0, \tag{19}$$

$$w(x,0) = w_0(x). \tag{20}$$

Here $w(x,t)$ is the vorticity, $u(x,t)$ is the velocity field, ν is the viscosity coefficient, and $f(x)$ is the forcing term. Let L denote the numerical solution operator mapping the initial vorticity w_0 to the final-time vorticity w_T. The observation model is

$$y = PL(w_0) + n, \tag{21}$$

where P samples the observed spatial locations and n denotes measurement noise.

This task is a nonlinear time-dependent fluid inverse problem whose forward map is defined by numerical PDE evolution. It is therefore a useful dataset for black-box ensemble guidance, since gradient through the full time-integration solver can be computationally burdensome. In UCEG, the observation units are spatial measurement locations of the vorticity field. We evaluate on 10 test cases.

B.3 DARCY FLOW PARAMETER ESTIMATION

We construct a Darcy flow parameter-estimation dataset on the unit square $\Omega = (0,1)^2$. The physical forward problem is defined by a transient Darcy/diffusion-type equation,

$$\partial_t u(x,t) - \nabla\cdot(a(x)\nabla u(x,t)) = f(x), \qquad x\in\Omega, \quad t\in[0,T], \tag{22}$$

with zero initial condition

$$u(x,0) = 0, \tag{23}$$

and homogeneous Dirichlet boundary condition

$$u(x,t) = 0, \qquad x\in\partial\Omega, \quad t\in[0,T]. \tag{24}$$

Here $a(x) > 0$ is the spatially varying coefficient field used in the PDE solve, $f(x)$ is the source field, and $u(x,t)$ is the PDE state.

The coefficient field is generated by a GRF/lognormal construction. We first sample a Gaussian random field

$$\tilde{g} \sim \mathcal{N}\left(0, (-\Delta + \tau^2 I)^{-2}\right), \tag{25}$$

which can be implemented by sampling discrete white noise ξ and solving the elliptic SPDE

$$(-\Delta + \tau^2 I)\tilde{g} = \xi. \tag{26}$$

The positive PDE coefficient is then obtained through the lognormal map

$$a(x) = \exp(\beta\tilde{g}(x)). \tag{27}$$

The source field is a localized Gaussian centered at a random location (x_c, y_c):

$$f(x) = A\exp\left(-\frac{(x-x_c)^2 + (y-y_c)^2}{\epsilon^2}\right). \tag{28}$$

The source center is sampled uniformly from an interior window,

$$(x_c, y_c) \sim \mathrm{Unif}([m, 1-m]^2), \tag{29}$$

where $m > 0$ keeps the source away from the boundary.

We discretize Eq. equation 22 using continuous Lagrange finite elements on a uniform triangular mesh. Time integration is performed with backward Euler. At each time step, the discretized system has the form

$$(M + \Delta t K(a))\, u^{n+1} = M u^n + \Delta t F, \tag{30}$$

Table 5: Main UCEG solver settings.

Task	Particles	Reverse steps	Guidance scale	Batch size
FWI	64	40	2.0	16
Navier–Stokes	2048	80	2.0	64
Darcy flow	256	80	0.20	32

where M is the mass matrix, $K(a)$ is the stiffness matrix assembled from the coefficient field $a(x)$, and F is the load vector corresponding to the source field $f(x)$.

Each raw sample contains the source field $f(x)$, the coefficient field $a(x)$ used in the PDE solve, the final-time PDE state $u(x, T)$, and the source location (x_c, y_c). All scalar fields are exported on a cell-centered $N \times N$ grid.

For training the joint FunDPS prior, we use a three-channel representation

$$(f, g, u), \tag{31}$$

where f is the source channel, g is the log-transformed parameter channel, and u is the log-transformed PDE-state / observation channel. In our preprocessing, all three channels are log-transformed before normalization. Thus, the target parameter channel g used in our reported Darcy reconstruction results corresponds to the logged parameter field derived from the coefficient used by the PDE solver.

We train the FunDPSYao et al. (2025) prior on 100,000 Darcy samples and evaluate inference on 20 held-out test cases. In UCEG, the observation units are spatial locations or spatial blocks on the state channel u.

C Experiment Details

This section reports the implementation details needed to reproduce the comparisons in Section 4. For FWI and Navier–Stokes, we follow the InverseBench protocols (Zheng et al., 2025b), including the pretrained plug-and-play diffusion priors used in that benchmark. Specifically, the FWI prior is trained on CurveFaultB subsurface velocity maps, and the Navier-Stokes prior is trained on vorticity fields obtained by evolving Gaussian-random-field initial vorticities for five time units. For Darcy flow, we use a SOTA prior FunDPS (Yao et al., 2025) trained on our Darcy dataset described in Appendix B.3. Unless otherwise stated, methods within each task use the same prior, observation model, and test cases.

UCEG solver settings. Table 5 summarizes the main UCEG solver settings. For FWI, unless otherwise specified, UCEG selects 8 sources out of 16, uses time stride 4, receiver budget 20 out of 129, uncertainty extra budget 5%, and effective uniform ratio 25%. Section 4.3 ablates the three receiver-side choices.

Baseline settings. For FWI, several derivative-free baselines require multiple hours for a single test case, so we use the strongest stable configurations available under our computational budget. For Navier-Stokes, we follow the InverseBench evaluation setup. Table 6 lists the method-specific settings. For FWI, all baselines use time stride 4 and receiver stride 1. DPS and SCG use a uniform 8-resources subset. For Navier–Stokes, we follow the corresponding InverseBench(Zheng et al., 2025b) settings for DPG, SCG, DPS-cGSG, and DPS-fGSG.

Darcy matched comparison. For Darcy flow, we use a controlled EnKG-vs.-UCEG comparison. Both methods use 256 particles, 80 reverse steps, two EnKG update rounds, guidance scale 0.20, and batch size 32. Table 7 lists only the observation-selection differences.

D Empirical Uncertainty Contraction Analysis

To further isolate the effect of the update-history monitor, we conduct a paired FWI trajectory experiment comparing UCEG with UCEG(random monitor). The two branches use the same

Table 6: Method-specific settings for FWI and Navier–Stokes.

Method	FWI				Navier–Stokes			
	Steps	Inner updates	Scale	Budget	Steps	Inner updates	Scale	Budget
EKI	–	12	2.0	32 particles	–	500	2.0	512 particles
EnKG	40	2	2.0	64 particles	80	2	2.0	2048 particles
DPG	24	–	10.0	64 samples	1000	–	180.0	5000 samples
DPS	24	–	0.1	8-shot subset	–	–	–	–
SCG	32	–	–	64 candidates	1000	–	–	512 candidates
DPS-cGSG	32	–	0.1	96 queries	1000	–	0.1	2048 queries
DPS-fGSG	32	–	0.1	96 queries	1000	–	0.1	2048 queries

Table 7: Darcy observation-selection settings.

Setting	EnKG	UCEG
Observation set	full grid	selected grid locations
Selection rule	full observation	update-activity monitor
Budget fraction	1.00	0.40
Effective uniform ratio	–	0.50
Uncertainty extra budget	–	2%

initial ensemble, first correction, shot subsets, uncertainty-guided receivers, diffusion configuration, correction schedule, guidance parameters, and total observation budget. They differ only in whether the monitor-controlled receivers are selected using the accumulated update history or replaced with random receivers.

For each correction step k, we measure ensemble uncertainty in the physical state space using the trace of the empirical covariance,

$$T_k = \mathrm{tr}\left(\widehat{\mathrm{Cov}}\left[\left\{x_{0,k}^{(j)}\right\}_{j=1}^{J}\right]\right). \tag{32}$$

Let T_k^- and T_k^+ denote the covariance traces immediately before and after the k-th EnKG correction, respectively. We define the relative contraction as

$$\rho_k = \frac{T_k^- - T_k^+}{T_k^-} \times 100\%. \tag{33}$$

For each test case, the mean relative contraction is obtained by averaging ρ_k over all correction steps. We additionally report the covariance trace after the final correction.

UCEG improves the mean relative covariance contraction by 0.512 percentage points, corresponding to a 2.61% relative improvement. The paired 95% case-cluster bootstrap confidence interval is $[0.283, 0.719]$ percentage points, and UCEG achieves stronger contraction in 9 of the 10 paired cases.

UCEG also reduces the final covariance trace by 28.03, corresponding to approximately 5.10%. The paired 95% confidence interval for the UCEG-minus-control difference is $[-43.43, -12.24]$, and UCEG again achieves a lower final covariance in 9 of the 10 cases.

These results do not imply that the realized update-activity map is an exact estimator of covariance contraction. Instead, they provide empirical evidence that the update-history ranking identifies observations that induce stronger uncertainty contraction under the practical nonlinear finite-ensemble algorithm.

E Additional Observation-Budget-Matched Controls

We provide additional controlled experiments to isolate the contribution of update-history-based observation allocation beyond the effect of observation reduction alone. In all comparisons below, UCEG(random) selects the complete reduced observation subset randomly, while UCEG(random

Table 8: Paired FWI analysis of empirical uncertainty contraction. UCEG and UCEG(random monitor) use the same observation budget and differ only in the receivers assigned by the update-history monitor.

Method	Mean relative contraction ↑	Final covariance trace ↓
UCEG(random monitor)	19.607%	549.19
UCEG	$\mathbf{20.119\%}$	$\mathbf{521.15}$

Table 9: Observation-budget-matched Navier–Stokes comparison under the $\times 8$, $\sigma = 2$ setting.

Method	Selected observations	Relative ℓ_2 ↓
EnKG	256	0.801110 ± 0.200196
UCEG(random)	64	0.605445 ± 0.225778
UCEG(random monitor)	64	0.569397 ± 0.211159
UCEG	64	$\mathbf{0.540658 \pm 0.207405}$

monitor) keeps the remaining UCEG pipeline unchanged and replaces only the observations assigned by the update-history monitor with random observations.

E.1 Navier–Stokes

We evaluate the additional controls under the $\times 8$ measurement subsampling setting with measurement noise $\sigma = 2$. All reduced-observation methods use the same 64-observation budget.

Under the identical 64-observation budget, UCEG reduces the relative ℓ_2 error by approximately 10.7% compared with UCEG(random) and by 5.0% compared with UCEG(random monitor). The latter comparison isolates the additional contribution of the update-history monitor because the remaining UCEG pipeline and observation budget are unchanged.

E.2 Darcy Flow

We further evaluate the same controls under a more selective 20% observation budget at ensemble sizes $N = 64$, 128, and 256.

UCEG achieves the lowest relative ℓ_2 error at all three ensemble sizes. The improvement over UCEG(random monitor) is approximately 1.2%–1.6%, while the improvement over UCEG(random) is approximately 1.5%–2.0%. Although the numerical differences are smaller than in FWI and Navier–Stokes, the same ordering is observed consistently across all three ensemble sizes.

F Additional Analysis of FWI Reconstruction Quality

We further examine whether the improved Relative ℓ_2, PSNR, and SSIM values of UCEG could primarily result from smoother reconstructions. To quantify recovery of spatial transitions and interfaces, we additionally evaluate the Sobel Gradient Relative ℓ_2 error. For both the reconstruction and ground truth, we compute horizontal and vertical gradients using normalized 3×3 Sobel filters, concatenate the two gradient components, and compute

$$E_{\mathrm{Sobel}} = \frac{\|\nabla_{\mathrm{Sobel}}\hat{x} - \nabla_{\mathrm{Sobel}}x\|_2}{\|\nabla_{\mathrm{Sobel}}x\|_2}. \tag{34}$$

Lower values indicate better agreement with the spatial transitions and interfaces of the ground-truth velocity field.

Table 11 reports the results on the same FWI cases for which complete paired full-observation EnKG reconstructions are available.

The Sobel Gradient Relative ℓ_2 error decreases from 0.585 for full-observation EnKG to 0.415 for UCEG with 10 corrections, corresponding to a 29.1% relative reduction. This indicates that

Table 10: Observation-budget-matched Darcy-flow comparison under a 20% observation budget.

	Relative ℓ_2 error ↓		
N	UCEG	UCEG(random monitor)	UCEG(random)
64	**0.21348**	0.21613	0.21682
128	**0.18327**	0.18592	0.18636
256	**0.16387**	0.16659	0.16717

Table 11: FWI reconstruction quality on paired cases. Sobel Gradient Relative ℓ_2 evaluates agreement of spatial gradients and interfaces.

Method	Relative ℓ_2 ↓	PSNR ↑	SSIM ↑	Sobel Grad. Rel. ℓ_2 ↓
Full-observation EnKG	0.0259 ± 0.0100	14.301 ± 2.974	0.398 ± 0.117	0.585 ± 0.121
UCEG, 10 corrections	$\mathbf{0.0210 \pm 0.0038}$	$\mathbf{15.562 \pm 1.537}$	$\mathbf{0.491 \pm 0.061}$	$\mathbf{0.415 \pm 0.042}$
UCEG, 20 corrections	0.0223 ± 0.0075	15.469 ± 3.081	0.449 ± 0.082	0.442 ± 0.066

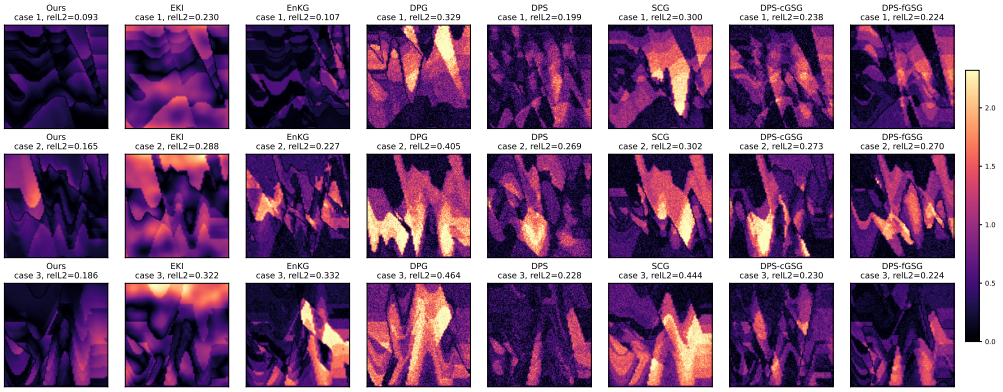

Figure 2: FWI absolute error maps corresponding to the reconstructions in Figure 1.

the improvement is also reflected in the spatial gradients of the reconstruction, rather than only in pixel-wise metrics that may favor smoother outputs.

We further increase the number of reduced-observation corrections from 10 to 20 while keeping the per-correction observation budget fixed. The 20-correction setting achieves a Relative ℓ_2 error of 0.0223, compared with 0.0210 for the 10-correction setting, while remaining better than full-observation EnKG across Relative ℓ_2, PSNR, SSIM, and Sobel Gradient Relative ℓ_2. These results indicate that the advantage of UCEG is maintained under a longer correction schedule, although increasing the number of corrections does not further improve the aggregate metrics.

G Additional Qualitative Visualizations

FWI error maps. Figure 2 provides the absolute error maps corresponding to the representative FWI reconstructions in Figure 1. Consistent with the quantitative results in Table 1, UCEG shows fewer large-error regions than the compared baselines, further supporting that the selected observation subset yields a reliable EnKG correction in this FWI setting.

Navier–Stokes reconstruction and error maps. Figures 3 and 4 provide additional qualitative results for the Navier-Stokes inverse problem under the $\times 2, \sigma = 0$ setting. Figure 3 compares the reconstructed vorticity fields, while Figure 4 shows the corresponding absolute error maps. The visual comparison is consistent with the quantitative results in Table 2: UCEG better recovers the target flow structures and reduces localized reconstruction errors compared with the baselines.

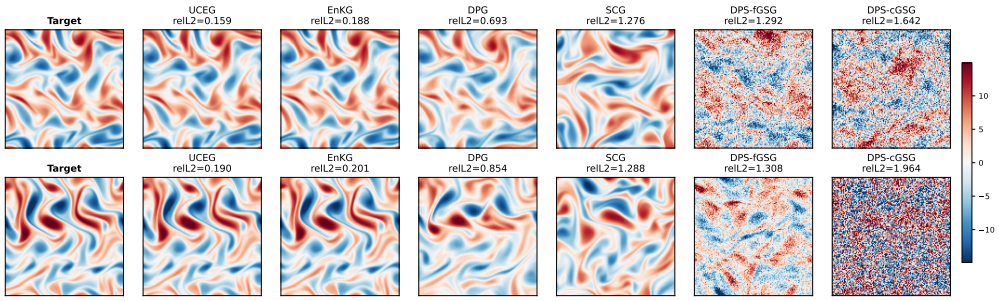

Figure 3: Navier-Stokes reconstruction comparisons under the $\times 2, \sigma = 0$ observation setting.

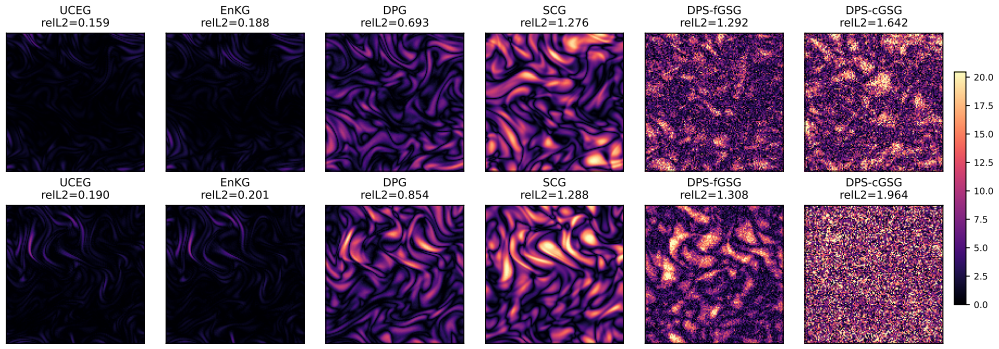

Figure 4: Navier-Stokes absolute error maps corresponding to Figure 3.

Darcy monitor-selection maps. Figure 5 visualizes the spatial regions selected by the UCEG monitor during Darcy inference. The figure includes the target field, the posterior-mean reconstruction, the cumulative selection map, and the per-update selection maps. The selections become increasingly concentrated in regions where the reconstruction remains inconsistent with the target field, especially in later updates. This behavior is consistent with the design of UCEG: regions that continue to induce stronger corrections receive higher monitor values and are therefore more likely to be selected in subsequent updates.

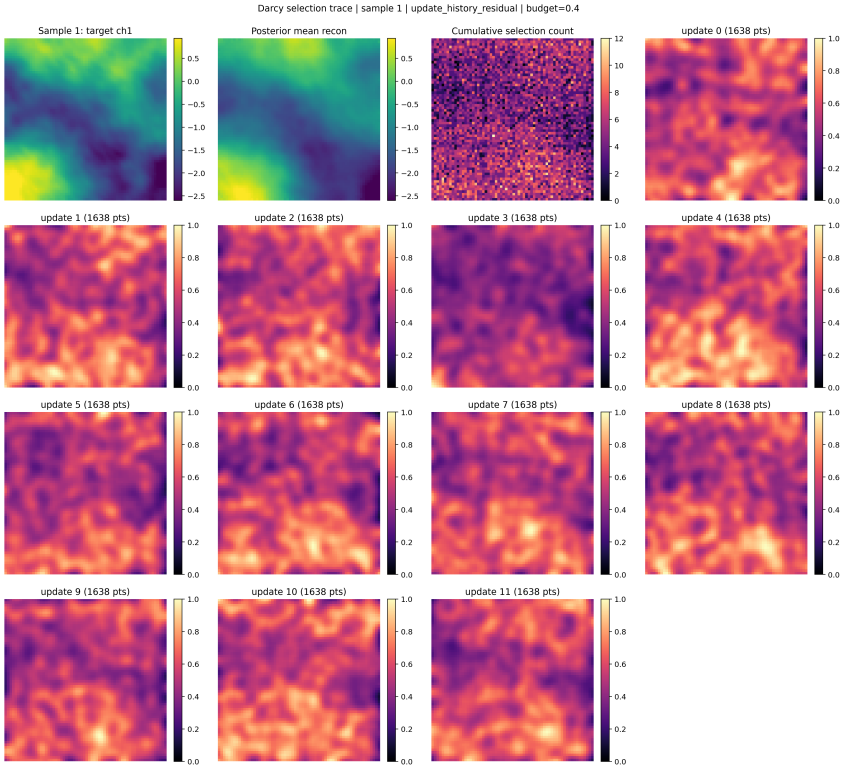

Figure 5: Darcy monitor-selection visualization. The panels show the target field, posterior-mean reconstruction, cumulative selection frequency, and per-update selection maps.